3a. "+-shift": in 1/4 of the cells the shift is 2*mu; in another 1/4 of the cells the shift is –2*mu; and in the remaining cells, the shift is zero. There are two variants:

3a1: the signs are spread evenly over the BST groups;

3a2: the + signs are concentrated in 1/4 of the groups, the – signs are concentrated in another 1/4 of the groups.

The plots show only the results for alternatives 1, 22, 32, 4, 3a1, 3a2. For the other two alternatives, namely 21 and 31, the "aveZ"[11] and "truncZ" curves are identical to those for 22 and 32; the BST curves are very similar to the "aveZ" curves. It should be noted that the "aveZ" curves are identical in each of the first four positions. This occurs because for this method the power depends only on the average shift. The power functions were computed setting the size of the test at 15%, 10%, and 5%. The only results that are shown are those occurring when the power function is computed setting the size of the test at 10%; however, the other two cases produce similar results.

As demonstrated in the first three panels, the power functions for all three procedures are very similar, except that "truncZ" is a little less powerful in the "allshift" and "halfshift" scenarios, while the BellSouth methodology ("BST") is a little less powerful when the deviations are concentrated in some of the groups rather than being spread over all. When only one mean is non-zero, "truncZ" is considerably more powerful than "aveZ", and "BST" is much less powerful than either; when the shift of the single nonzero mean is large, the "BST" procedure has power < size. This is because

[11] The term "aveZ" is a reference to the Adjusted z2 test.

the large shift contributes to the variance in the denominator of the BST statistic, while the numerator is small; thus the variance of the "BST" statistic is smaller than under the null hypothesis, and it is less likely that the statistic will exceed the critical value. When there are deviations of both signs, these deviations averaging to zero, and being spread over the BST groups, both "BST" and "aveZ" have power = size, so these tests are completely ineffective. The "truncZ" statistic does have some chance of detecting this alternative. If the positive deviations are concentrated among the BST groups, the "BST" procedure has power < size; again, this is because the differences contribute to the variance in the denominator of the BST statistic, making it less likely that the statistic will exceed the critical value.

3. How important is it to balance the probability of Type I and Type II errors? Is there a mechanical formula that would adjust the critical values (and hence the probability of a Type I error) as the sample size varied? How can we explicitly measure the costs of a Type I and of a Type II error, as BellSouth suggests needs to be done?

Response to Question No. 3:

Importance of Balancing

In any statistical analysis, there are inherent risks of reaching one of two distinct types of testing errors. "Type I" errors occur when a statistical test reveals that the ILEC is not meeting its obligation to provide parity of service when, in fact, it is. BellSouth, of course, would like to minimize the probability of Type I errors. By contrast, Type II errors occur when a statistical test reveals that the ILEC is providing parity of access when, in fact, it is not. From the CLEC's perspective, the statistical test procedure should be designed so as to minimize the probability of Type II errors.

Both types of errors are possible and important in determining whether the null hypothesis should be accepted. Type II errors are as real as Type I errors. As a result, there may be instances in which the ILEC is not providing equal service to CLECs, however, purely by chance the statistical test fails to reject the null hypothesis. As a result, it is necessary to strike a balance between the two types of errors. If the Type I error rate selected in the statistical methodology is too small, the Type II error rate will be large. The converse is also true.

Critical Values

There is an algorithm that can generate the appropriate value of the Type I error so that it equals the Type II risk. The algorithm requires the specification of the distribution of the chosen test statistic under both the null and some alternative hypotheses. These will depend on the observed two sample sizes. Given the two sample sizes, it is easy to determine the critical value so that the Type I and Type 2 errors are equal. For example, if the two distributions are standard normal and shifted normal and if the sample sizes are equal, then the critical value is halfway between the null mean and the alternative mean. The algorithm can easily be expressed in computer code.

The So-Called Measure of Costs

AT&T and BellSouth agree that the probability of Type I and Type II errors should be balanced, and that the balance of the probabilities of Type I and Type II errors should depend upon: "1. the effective number of BellSouth observations; 2. [t]he effective number of CLEC observations; [and] 3. [t]he size of a specific alternative hypothesis, e.g. the CLEC mean value is larger than the BellSouth mean value by ten percent of a BellSouth standard deviation." See Statistical Procedure Off-Line Session; Consensus/Open Issues appended to Letter from Kathleen Levitz to the Hon. Magalie Roman Salas dated May 20, 1999 (CC Docket Nos. 98-56, 98-121). However, BellSouth proposes to use an alternative procedure if the balancing procedure is somehow unworkable. Under BellSouth's alternative procedure, the size of a difference between mean values which has no business impact must be defined. BellSouth refers to this procedure as the rule of materiality. Any actual difference that is considered less than the materiality standard will be considered insignificant; conversely, any difference that is greater than the materiality standard will be deemed significant based on a statistical testing procedure. In addition, BellSouth has proposed using the conventional level of 5%, a two-sided equivalent to a 2.5% one-sided Type I error.

AT&T does not agree that the costs of a Type I and of a Type II error can be explicitly measured. For the ILEC, the cost of a Type I error -- an error that can be treated as the cost of doing business -- can be neutralized through provision within a system of penalties and remedies. By contrast, for the CLEC, a Type II error could not only result in the ILEC's successful avoidance of detection of discrimination, but could also delay or thwart entirely the CLEC's competitive entry into the local market. Under such circumstances, the cost of a Type II error could be incalculable, but enormous.

Further, BellSouth's proposed materiality standard is fundamentally infirm. Even assuming arguendo that the parties could somehow agree on a so-called materiality standard, BellSouth proposes to abandon an approach requiring equalization of the risks of Type I and Type II errors and to substitute therefor an arbitrary 2.5% significance level that is biased in BellSouth's favor for all sample sizes below 1000.

Indeed, under BellSouth's proposal, although there is a smaller risk of a Type I error occurring (a smaller risk of declaring BellSouth to be out of parity when it is really in parity), there is an increased risk of a Type II error (not declaring BellSouth to be out of parity when it is).

Fundamental fairness requires a statistical approach that treats each carrier equally. The only fair and rational basis for determining how low the risk of false accusation should be is to equalize the risks borne by the ILEC and CLEC of any error counter to its interests. For purposes of testing the results of individual performance measures, the selection of the compromise value of the Type I error should depend on the CLEC sample size. Thus, for example, the following values that are tied to CLEC sample size could be used:

CLEC Sample Size	alpha=beta risk
10	.2564
30	.2538
60	.2459
100	.2337
300	.1731
600	.1068
1000	.0525

These values are based upon the following assumptions: the populations are normal; the BellSouth sample is much larger than the CLEC sample; and the alternative that is to be detected is a shift of the population by one-tenth of a standard deviation.

Concerning Estimating the Variance:

4. Why is it desirable to use replication to estimate the variance? What advantages does this have over using an alternative method?

Response to Question No. 4:

AT&T does not support the replication method to estimate the variance. The disadvantage of using replication to estimate the variance is that it allows variance between cells to contribute to the estimate of the overall variance. As a consequence, if there are systematic differences between cells, it is less likely that discriminatory performance will be detected. The power calculations set forth in response to Question No. 2 above show that when there are either shifts concentrated in just one cell, or shifts of opposite signs that cancel out in the numerator but inflate the variance in the denominator, BellSouth's methodology is completely ineffective in detecting discriminatory performance.

Concerning Aggregating the Data:

5. What are the specific advantages/disadvantages of using aggregation of the adjusted data (the BellSouth approach)? Compare to testing unadjusted aggregate data (LCUG's original proposal) and testing individual cells of disaggregated data (LCUG's recent approach)? In particular, consider the criteria discussed in question 1.

Response to Question No. 5:

As noted above, LCUG has not endorsed a test involving the calculation of the LCUG statistic to unadjusted aggregate data. In addition, as AT&T noted above, a statistical methodology that uses unadjusted aggregate data is not a valid test of discriminatory performance. Further, LCUG Version 1.0 -- the only statistical test adopted by LCUG -- does not address the level of disaggregation that is required so that "like-to-like" comparisons of CLEC and ILEC data can be made. In order to test whether an ILEC has provided and is providing discriminatory service, it is important to disaggregate observations in each service quality measurement into cells that are small enough so that "like-to-like" comparisons can be made. AT&T states, by way of further answer, that its response to Question No. 1 herein contains additional information that is responsive to Question No. 5.

6. Are there tests that can be performed to determine the validity of the degree of aggregation that BellSouth proposes versus the degree of disaggregation LCUG proposes? Is there some middle ground that can be reached through such tests by aggregating some of the cells, where appropriate, and disaggregating where aggregation is not appropriate?

Response to Question No. 6:

As noted above, LCUG did not propose any level of disaggregation in LCUG Version 1.0. While there are tests that can be performed to determine, for example, the validity of the degree of aggregation that BellSouth proposes versus the level of disaggregation proposed under the truncated method, such tests would be unduly complicated. In this regard, AT&T agrees with BellSouth that the "[a]ppropriate 'middle ground' would change from month to month [and would not be] feasible or consistent with black box/production mode." See Chart titled "Middle Ground" attached to 5/20/99 Levitz Letter.

Concerning Dependency:

7. Isn't the replicate estimate of the variance also affected by dependency (i.e., correlation) in the data? This appears to be confirmed in Wolters. Which methodology is affected more by dependency in the data?

Response to Question No. 7:

As its response to Question No. 7, AT&T incorporates by reference its response to Question No. 1 herein.

8. How much dependency is there in the data (between measures, wire-centers, over time)? How can this be determined? Should this be determined using statistical means, or by examining physical relationships involved between measures (dependence on common computer system or common cable), or by examining each event *ex post*? Can a covariance matrix be developed using weekly or daily data, or by matching the ILEC and CLEC data using the multiple cells created through disaggregation? How much will the dependency measured affect the probability of a Type I error for the LCUG method, and for the BellSouth method?

Response to Question No. 8:

In order to respond to Question No. 8, AT&T must have unfettered access to BellSouth's performance data. In this regard, pursuant to a Protective Agreement entered into between BellSouth and AT&T on February 9, 1999, Dr. Colin Mallows was permitted to examine BellSouth's proprietary performance data. However, the Protective Agreement provides, inter alia, that "[a]ny information obtained by Dr. Mallows from his examination of the proprietary and confidential information will be used solely in connection with LPSC Docket No. U-22252-C or any appeal of a decision of the Commission arising from that docket and for no other purpose or in connection with any other docket or proceeding." Protective Agreement, In Re: BellSouth Telecommunications, Inc. Service Quality Performance Measurements, Docket U-

22252, February 9, 1999. The Protective Agreement further provides that Dr. Mallows is precluded from divulging "information obtained from an examination of the confidential data" to any person. Because the Protective Agreement precludes Dr. Colin Mallows from disclosing BellSouth's proprietary data and any information derived therefrom to any other person or in other proceedings, it is impossible for AT&T to respond to Question No. 8. AT&T and BellSouth have engaged and continue to engage in discussions regarding the contours of a procedure under which Dr. Mallows' analysis of BellSouth's data can be released to the FCC. If the parties reach agreement regarding such matters, AT&T will supplement its response to Question No. 8.

Concerning Normality of the Data, and Sample Sizes:

9. Are the data nonnormal? How can this be determined? What size sample do we need to get an approximately normal distribution of a mean? How can this be determined?

Response to Question No. 9:

In order to respond to Question No. 9, AT&T must have unrestricted access to BellSouth's data. As AT&T noted in its response to Question No. 8, the Protective Agreement between AT&T and BellSouth precludes Dr. Mallows from divulging BellSouth's proprietary data or any information derived therefrom to any other person or in other proceedings. If BellSouth and AT&T are able to reach agreement regarding the release of Dr. Mallows' analysis of BellSouth's data to the FCC, AT&T will supplement its response to Question No. 9.

10. Is permutations testing the best way to handle small, nonnormal samples? What are the advantages and disadvantages of permutations testing? Are there any problems that small sample sizes create for BellSouth's proposed methodology? Could we see a comparison of the results using permutation testing with BellSouth's results?

Response to Question No. 10:

Permutation testing is the best way to handle small, nonnormal samples. Under the permutation test, the probability distribution is generated through use of the actual sample results. Two samples (X's and Y's from ILEC and CLEC, respectively) can be combined into one pool and then divided into two sets X* and Y* in all possible ways. For each way, the corresponding z-score (z*) can be calculated. This yields a distribution of z* values, each of which is equally likely under the null hypothesis that the ILEC is treating customers impartially. Given the desired Type I error rate, one can read off the appropriate critical value and compare this with the observed value. For example, assume the data are the following:

3 ILEC observations: X=1, X=2, X=4
2 CLEC observations: Y=3 and Y=5

Accordingly, the pooled set is (1,2,3,4,5), and there are 10 ways one can assign these five observations to the ILEC and CLEC samples. 10 values of z can be calculated (-2.74 -1.20 -0.6 –0.44 0.00 0.00 0.44 0.60 1.20 2.74), and the 5% critical value is 2.74. The actual observed value is 1.20, and so is judged to be not significant.

This test procedure is valid irrespective of the form of the population distribution, since it depends only on the assumption that each possible permutation is equally likely under the null hypothesis. The method can be used whenever the sample sizes are large enough to make the test statistic well defined, in the present case even for m=1, n=1. Of course, with such small samples, only very coarse results are obtained, but results from a large number of small samples can be aggregated to generate meaningful results.

The permutation distribution would be developed through the use of a computer program enumerating the samples necessary to generate the distribution. A resource issue relating to the use of the permutation distribution is the time needed to generate results. Unless the sample sizes are very small, the number of permutations to be generated is extremely large. In order to deal with this problem, it would be reasonable to use a random sample of possible permutations to approximate the distribution. For example, if the number of possible permutations in a particular case exceeds 1000, the program could be designed to approximate the permutation probability distribution by randomly selecting 1000 permutations and constructing the distribution from those data. Because computers can perform calculations such as this with remarkable speed, the distribution for any measurement category could be ascertained within a few seconds.

BellSouth's methodology can handle small samples. The FCC has requested that AT&T provide "a comparison of the results using permutation testing with BellSouth's results." Because BellSouth's data are proprietary and because the Protective Agreement between AT&T and BellSouth precludes Dr. Colin Mallows from releasing BellSouth's proprietary performance data or any information derived therefrom to any person or in other proceedings, it is impossible for AT&T to respond to the FCC's request for an analysis of permutation testing using BellSouth's performance results. If AT&T and BellSouth are able to reach agreement regarding the release of Dr. Mallows' analysis of BellSouth's performance data to the FCC, AT&T will supplement its response to Question No. 10.

Concerning Statistical vs. Competitive Significance of the Results:

11. Should a statistically significant difference in means be interpreted to mean that there is discrimination in the process? In other words, should we consider whether the observed difference in means will have an economic impact on CLECs' business? Won't very large sample sizes tend to make even small differences in means statistically significant? How large should a difference in means be for a particular measure for it to be considered "competitively significant" and therefore discriminatory? How should this "threshold difference" be determined for each measure? How can a "threshold difference" be implemented for a testing procedure?

Response to Question No. 11:

The FCC has found that the term "nondiscriminatory" as used in the Telecommunications Act of 1996 (the "Act") is a more stringent standard than the "unjust and unreasonable discrimination" standard set forth in other provisions of the Communications Act.[12] Thus, the "nondiscriminatory access" a BOC must provide under the competitive checklist means that: (1) the quality of the interconnection or UNEs that the BOC provides, as well as the access to such interconnection or UNES, must be equal among all carriers requesting access; and (2) where technically feasible,

[12] See FCC Docket No. 96-98, Implementation of the Local Competition Provisions in the Telecommunications Act of 1996, First Report and Order released August 8, 1996, ¶¶ 217, 859 ("Local Competition Order").

the access and interconnection or UNES must be at least "equal" in quality to that which the BOC provides to itself.[13]

As part of this nondiscrimination requirement, the FCC has made clear that a BOC must monitor its performance in accordance with an appropriate measurement plan that produces empirical data that allows its performance to be evaluated fully and accurately. As the FCC has observed, "proper performance measures with which to compare BOC retail and wholesale performance, and to measure exclusively wholesale performance, are a necessary prerequisite to demonstrating compliance with the Commission's 'nondiscrimination' and 'meaningful opportunity to compete standards.'" Ameritech Michigan Order, ¶ 204 (quoting DOJ Ameritech Michigan Evaluation, filed June 25, 1997, App. A, p. A-3). The FCC has also emphasized that, in order to demonstrate that nondiscriminatory access is being delivered to CLECs, a BOC must provide empirical data showing that the access is "the same" as or "equal to the level of access that the BOC provides to itself, its customers,

[13] Local Competition Order, ¶ 315 (access must be provided on terms that are "equal to the terms and conditions under which the incumbent LEC provisions such elements to itself"); Second Order on Reconsideration, Implementation of the Local Competition Provisions in the Telecommunications Act of 1996, CC Docket No. 96-98 (released December 13, 1996) ¶ 9 (OSS access "must be equal to" the access that the ILEC provides to itself); FCC CC Docket No. 97-137, In the Matter of Ameritech Michigan Pursuant to Section 271 of the Communications Act of 1934, as amended, To Provide In-Region InterLATA Services in Michigan, Memorandum Opinion and Order released Aug. 19, 1997 ("Ameritech Michigan Order"), ¶ 139 ("BOC must provide access to competing carriers that is equal to the level of access that the BOC provides to itself . . . in terms of quality, accuracy and timeliness"); ¶ 166 (ILEC "must provide to competing carriers access to such OSS functions equal to the access that it provides to its retail operations").

or its affiliates, in terms of quality, accuracy and timeliness." Ameritech Michigan Order, ¶ 139.

In addition, the FCC has expressly held that "Section 271 places on the applicant the burden of proving that all of the requirements for authorization to provide in-region, interLATA services are satisfied. . . . [T]he ultimate burden of proof with respect to factual issues remains at all times with the BOC, even if no party opposes the BOC's application." Ameritech Michigan Order, ¶ 43 (emphasis added; footnotes omitted). See also BellSouth South Carolina Order, ¶ 37. Thus, in the first instance, a BOC must present a prima facie case in its application that all of the requirements of Section 271 have been satisfied. Only if the BOC's application makes such a case are opponents of the BOC's entry required, "as a practical matter," to produce evidence and arguments necessary to show that the BOC does not comply with the checklist. However, "the BOC applicant retains at all times the ultimate burden of proof that its application satisfies section 271." Ameritech Michigan Order, ¶ 44 (emphasis added).[14]

It is true that, as the number of data points underlying the computed performance measurement increases (all other factors held constant), smaller differences in the mean will be statistically significant. However, if the risks of Type I and Type II errors are appropriately balanced, the probability of Type I and Type II

[14] See also BellSouth South Carolina Order, ¶¶ 37, 57. Although the July 8 Order (which was issued prior to the FCC's Ameritech Michigan Order) refers frequently to BA-NY's failure to make a prima facie case, the FCC's subsequent orders have made clear that the BOC not only must make a prima facie case of compliance with the checklist but also bears the burden of proof on all issues in the Section 271 proceeding. See Ameritech Michigan Order, ¶¶ 43-44; BellSouth South Carolina Order, ¶ 37; July 8 Order, pp. 1, 4, 12-13, 19-20, 22, 24, 33.

errors will be small with large sample sizes. Under such circumstances, a statistically significant difference in means evidences discriminatory performance.

It must also be emphasized that the term "discrimination" under the Act should not be confused with competitive injury. In that connection, the Act nowhere states that a BOC that discriminates against a CLEC can, nonetheless, satisfy its statutory obligations under Section 271 by showing that its discriminatory performance failed to meet some threshold for competitive injury sustained by the CLEC. As a theoretical proposition, it is impossible to state how large a difference in means should be for a given performance metric before it should be considered "competitively significant." Classical economic theory would suggest that any discriminatory performance will increase the CLEC's costs and/or degrade service and thereby impair the CLEC's ability to compete in the marketplace. A more precise answer would require extensive analysis of, inter alia, the carrier, the entry strategy, the volume of business affected, and the particular measurement in question.

In any event, it must be emphasized that the burden for demonstrating non-discrimination rests with the BOC. Ameritech Michigan Order, ¶ 43. Therefore, if a statistically significant difference in means exists, the BOC must bear the burden of showing that it has satisfied its obligations under the Act. To do otherwise would shift the burden of proof to the CLEC -- a perverse outcome that is inconsistent with the FCC's decisions.

Concerning Confounding Factors:

12. Is it necessary to disaggregate according to every confounding factor? What are the advantages and disadvantages of doing so? Would it be possible to disaggregate only according to those confounding factors that are statistically determined to have an impact on the results?

Response to Question No. 12:

Performance results must be "sufficiently disaggregated to permit meaningful comparisons." Ameritech Michigan Order, ¶¶ 212, 206.[15] Accordingly, an effective performance measurement plan must collect and mark the data used to calculate the measurements so that meaningful "apples-to-apples" comparisons of results for the CLECs and ILEC can be made. Thus, for example, in commenting on Ameritech's Section 271 application, the Michigan Public Service Commission observed:

> Measurements must be refined enough to permit meaningful parity comparisons to be made. That is, if business orders are more complex and handled differently by Ameritech's retail operations than are residential orders, performance measures should distinguish these operations. Separate measurements for different customer classes, geographic areas or service products may be required.

Michigan Public Service Commission Comments, FCC Docket No. CC 97-137, filed June 9, 1997, pp. 31-32.

---

[15]. See also In the Matter of Application of BellSouth Corporation, BellSouth Telecommunications, Inc. and BellSouth Long Distance, Inc., for Provision of In-Region, InterLATA Services in Louisiana, FCC Memorandum and Opinion issued October 13, 1998 ("BellSouth Second Louisiana Order"), ¶ 92.

Similarly, a meaningful "apples-to-apples" comparison requires that performance data for both CLECs and ILEC be reported for the same geographical market area. As the Department of Justice has explained:

> Geographic parity requires that performance measures be identified and measured where a CLEC markets [its] products. . . . If a CLEC offers service to smaller geographic areas, appropriate performance measures would provide comparative BOC results for those areas.

Affidavit of Michael J. Friduss on behalf of the Department of Justice, filed November 4, 1997, in Application by BellSouth Corp., et al. for Provision of In-Region, InterLATA Services in South Carolina, CC Docket No. 97-231, ¶ 68.

The need for disaggregation can arise because of differences in geographical location, differences in products or services, or differences in the tasks which must be performed by the ILEC, such as installations or repairs that can be accomplished at the central office versus installations and repairs that require that a technician be dispatched to the customer's premises. For example, an order requiring only a software change can be completed in less than a day, while an order requiring a new drop wire may take several days. Similarly, different maintenance situations – such as service affecting or out-of-service – may have different priorities and, accordingly, different repair intervals for the same service. If these different services or activities are all combined into a single measure, the results are likely to be misleading, particularly because the ILEC and the CLECs are likely to have different product and service mixes.

An appropriate level of disaggregation of performance results should not be unreasonably burdensome for the ILEC. The ILEC will collect the same or similarly structured data in any event. Appropriate marking and categorization of the data as it is collected are all that is required. The associated measurements are still computed in the same manner that they would be without the disaggregation requirement. The benefits of disaggregation in encouraging and protecting competition should more than compensate for any additional costs.

BellSouth appears to advocate disaggregation by wire center, time of month, and at those levels identified by the Louisiana Public Service Commission. By contrast, AT&T advocates using the levels of disaggregation proposed by LCUG in SQM Version 7.0. See Exhibit 3 (setting forth levels of disaggregation applicable to product and activity). There are statistical regressions that can be used to determine whether a particular variable is correlated with differences in performance. However, only BellSouth possesses the performance data required to assess the precise level of disaggregation that would be required to provide an accurate assessment of its performance. In any event, disaggregation should be at a level where relatively few expected dissimilarities in performance exist, so that both the mean or average performance of the group and the expected variance should be the same.

This Appendix discusses the calculation of the mean and variance of the proposed overall statistic Z_{agg}. We have

$$M = \text{sum}(W_i E(Z_i^*)), \qquad V = \text{sum}(W_i^2 \,\text{Var}(Z_i^*))$$

If the sample sizes are large, then to a good approximation Z has a standard Normal distribution and

$$E(Z^+) = 1/\sqrt{(2\pi)} = .3989, \qquad Var(Z^+) = 1/2 - 1/2\pi = .3408$$

However, if the sample sizes are small and Z is obtained by a permutation calculation, then Z will be only approximately Normal. For example, if $^n{}_{ILEC}$ = $^n{}_{CLEC}$=2 and parity holds, and if the observations are in general position so that no ties occur, then the distribution of the permutation Z is uniform on the six values

Q(1/12)	Q(3/13)	Q(5/12)	Q(7/12)	Q(9/12)	Q(11/12)

where Q is the inverse Normal function; these values are

-1.3830	-0.6745	-0.2104	0.2104	0.6745	1.3830

Thus Z^+ is uniform on the six values

-1.3830	-0.6745	-0.2104	0	0	0 and

$E(Z^+)$=-0.3780, $\qquad$ $Var(Z^+)$=.2591

We can correct for this effect by calculating $E(Z^+)$ and $Var(Z^+)$ directly, as functions of $^n{}_{ILEC}$ and $^n{}_{CLEC}$. This calculation is not onerous, and can easily take account of ties.

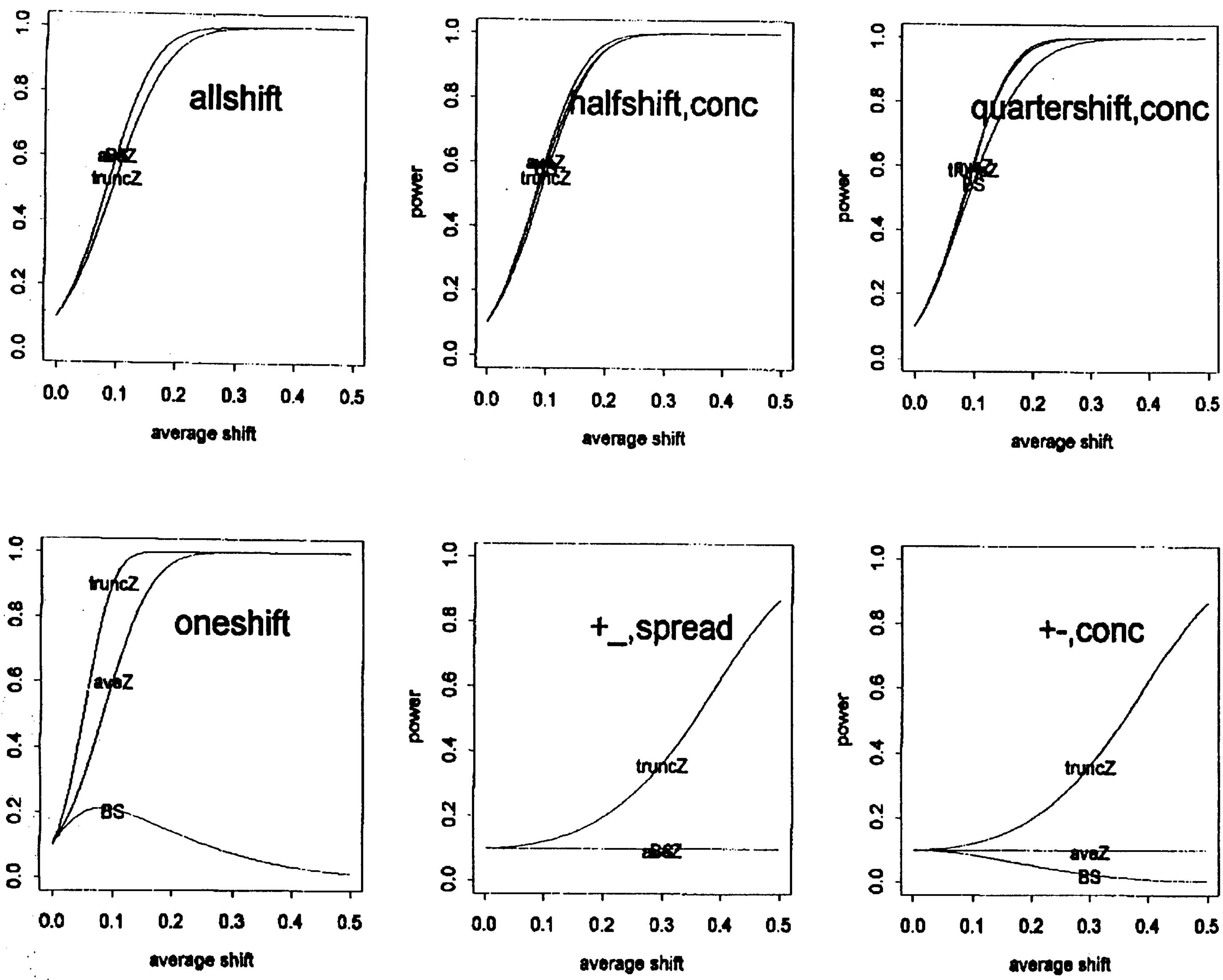

allshift
truncZ
average shift
power
halfshift,conc
truncZ
quartershift,conc
BS
oneshift
truncZ
avaZ
BS
+_,spread
truncZ
+-,conc
truncZ
avaZ
BS

EXHIBIT 2

Here follow the S functions that generated the power functions and the plots.

```
getW_function(n=8){   # generates a matrix W such that WX gives the group means.
d_rep(c(rep(1,n),rep(0,30*n)),30)
t(matrix(d[1:(900*n)],30*n,30))/n
}

tmean_function(x)dnorm(x)+x*pnorm(x)          # finds the mean of trunc(mu+X)
tvar_function(x)x*tmean(x)+pnorm(x)-tmean(x)^2  # finds the variance of trunc(mu+X)

mus_(0:50)/100
zs_c(qnorm(.85),qnorm(.90),qnorm(.95))

getpowers_function(mus,zs,n=8,W=getW(n)){       # assumes 30 groups of size n
nmus_length(mus)
nzs_length(zs)
N_30*n
sqN_sqrt(N)
a_dnorm(0)                                 # mean of trunc(X)
s_sqrt(.5-a^2)                             # s.d. of trunc(X)
ans_array(0,dim=c(nmus,nzs,8,3))
for(i in 1:nmus){
mu_mus[i]
mu1_rep(mu,N)                              # hypothesis 1
mu21_rep(c(rep(0,n/2),rep(2*mu,n/2)),30)   # hypothesis 2, spread over the groups
mu22_c(rep(0,N/2),rep(2*mu,N/2))           # hypothesis 2, concentrated
mu31_rep(c(rep(0,3*n/4),rep(4*mu,n/4)),30) # hypothesis 3, spread over the groups
mu32_c(rep(0,3*N/4),rep(4*mu,N/4))         # hypothesis 3, concentrated
mu4_c(N*mu,rep(0,N-1))                     # hypothesis 4
mu3a1_rep(c(rep(2*mu,n/4),rep(-2*mu,n/4),rep(0,n/2)),30)   # hypothesis 3a, spread
mu3a2_c(rep(2*mu,N/4),rep(-2*mu,N/4),rep(0,N/2))          # hypothesis 3a, concentrated
mumtx_cbind(mu1,mu21,mu22,mu31,mu32,mu4,mu3a1,mu3a2)   #put all these into a matrix
gmeans_W%*%mumtx                                       # means ofthe BST groups
tm_tmean(mumtx)                                        # means of truncated X's
tv_tvar(mumtx)                                         # variances of truncated X's
for(j in 1:nzs){
for(k in 1:8){
mk_gmeans[,k]                                                    # the group means
ans[i,j,k,1]_pnorm(sqN*mean(mk)-zs[j]*sqrt(1+n*var(mk)))         # "BST" method
ans[i,j,k,2]_pnorm(sqN*mean(mk)-zs[j])                           # "aveZ" method
ans[i,j,k,3]_pnorm((sum(tm[,k])-N*a-zs[j]*sqN*s)/sqrt(sum(tv[,k]))) # "truncZ" method
}}}
ans
}

powers_getpowers(mus,zs)                   # run the function

par(mfrow=c(2,3))                          # six plots on the page
Hyp_c("allshift","halfshift","halfshift,conc","quartershift","quartershift.conc","oneshift",
"+_,spread","+-,conc")                         #
wherex_c(.1,.1,.1,.1,.1,.1,.3,.3)               #
wherey_c(11,11,11,11,11,11,31,31)               # labels for the plots
Method_c("BS","aveZ","truncZ")                  #
for(j in 2){                               # make the plot
for(k in c(1,3,5,6,7,8)){
plot(mus,powers[,j,k,1],type="l",xlab="average shift",ylab="power",ylim=c(0,1))
text(.3,.8,labels=Hyp[k],cex=1.2)
text(wherex[k],powers[wherey[k],j,k,1],labels=Method[1])
for(m in 2:3){
lines(mus,powers[,j,k,m])
text(wherex[k],powers[wherey[k],j,k,m],labels=Method[m])
}}}
```

A comment on the BellSouth simulation.

The four attached sheets relate to a simplified version of the BellSouth simulation. The simplifications are:

(i) the sample sizes are held constant, nILEC = 100, nCLEC = 5 in each wire-center ("w/c");

(ii) the within-w/c correlations are all 0.35;

(iii) there is only one subclass within each w/c;

(iv) there are no w/c mean or standard deviation effects.

The first two pages show two runs of the simulation. The effect of the presence of the within-w/c correlation is, for the data for each company,

(a) to multiply the within-w/c variance by a factor 1-0.35 =0.65,

(b) to add random shifts with variance 0.35 to all the observations within each w/c.

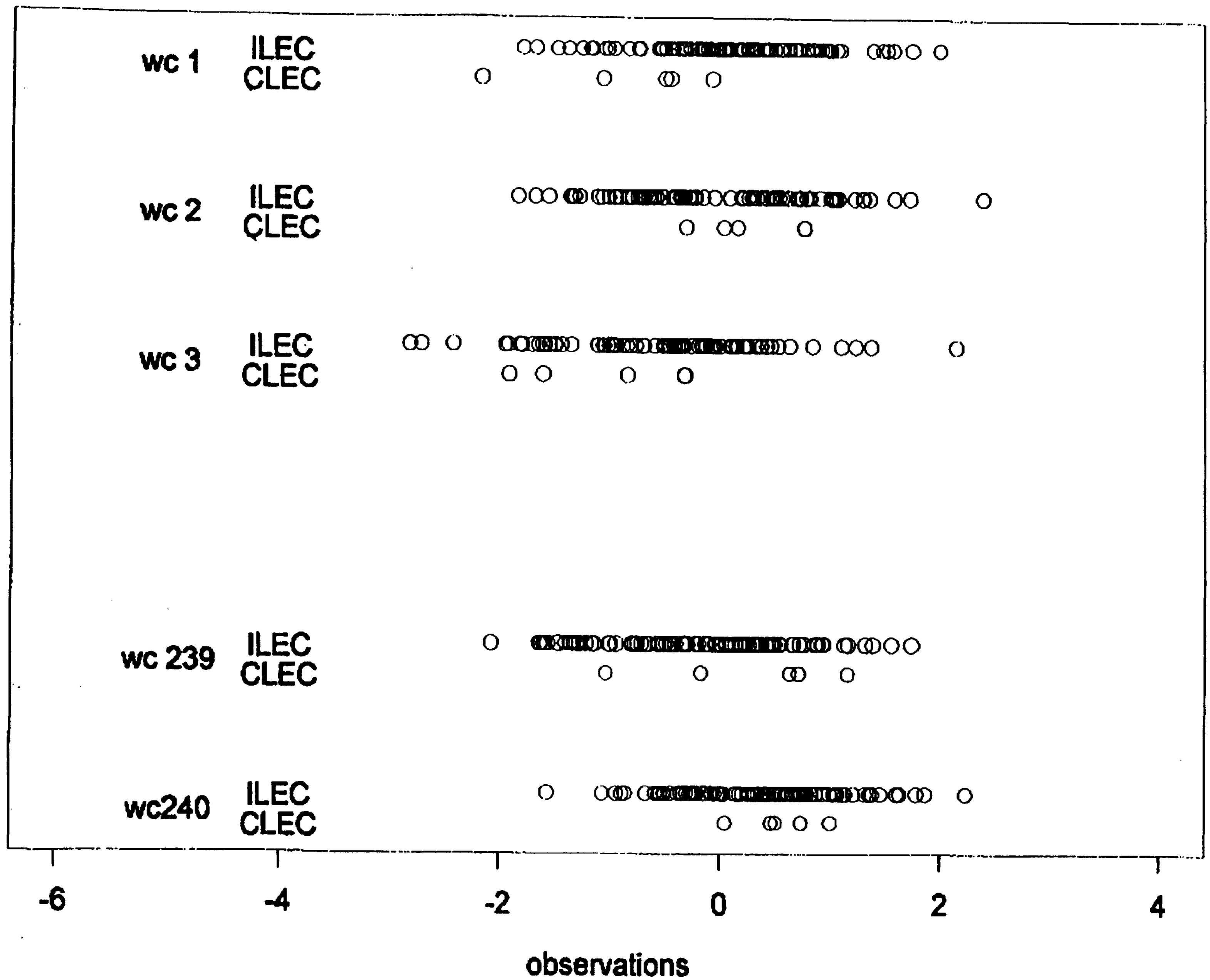
In each w/c, there are 100 ILEC observations and 5 CLEC observations. The sample means vary randomly (and independently) about zero. The overall variance, for each company and each w/c, is unity. In successive simulation runs, everything is independent of the previous runs.

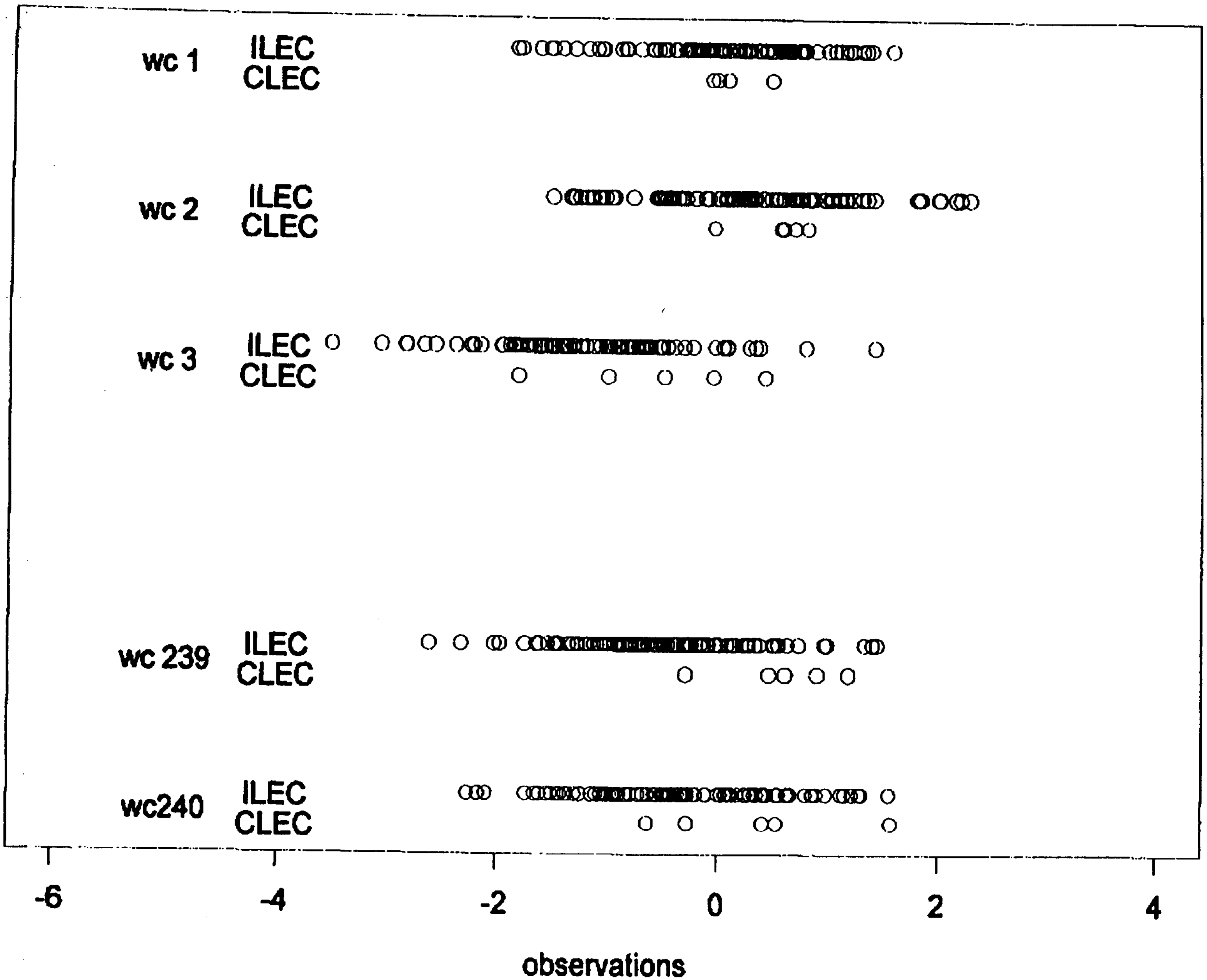
In generating the third and fourth sheets, a very small change to the simulation program was made. After generating all the sample data, the w/cs were sorted according to the values of the difference between the two random shifts that occurred for that w/c. Thus, w/c #1 gets the largest value of {shift(CLEC – shift(ILEC)}, w/c #2 has the second largest value, and so on; w/c #240 gets the smallest (negative) value of this difference. This is done for every simulation run. The only difference between the first and third sheets lies in the labels attached to the w/cs, and similarly for the second and fourth sheets. This change will make no difference to the BellSouth method.

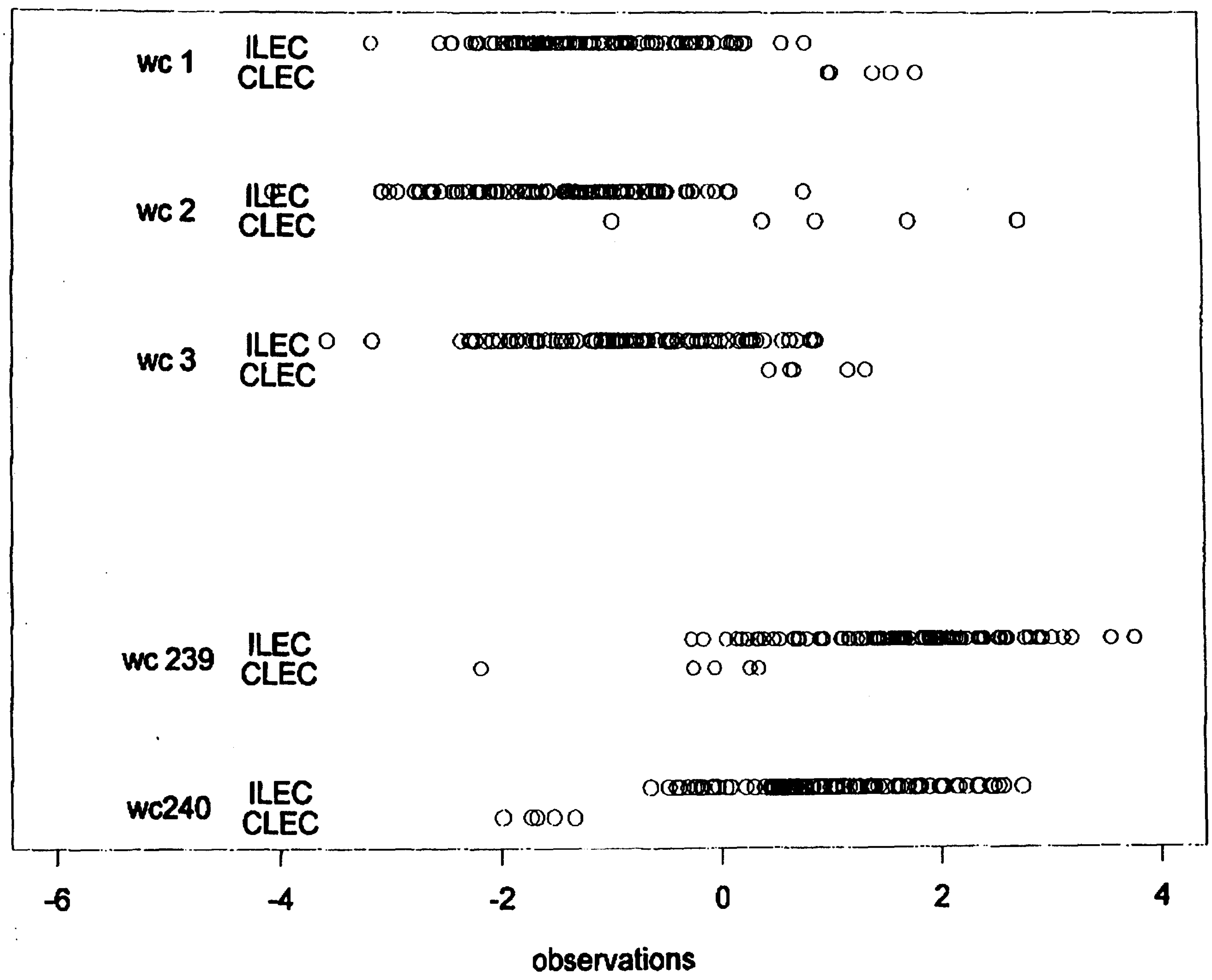
After making this change, the effect of the within-w/c correlation can be reinterpreted. In w/c #1, in each simulation run, the difference (ILEC mean – CLEC mean) is about –1.7; in w/c #2 this difference is about –1.5, and so on. If data like this were to occur in the real world, these effects would be interpreted as being consistent biases associated with the w/cs. In w/cs #1-120, there is consistent discrimination against the CLEC, while in w/cs 121-240 the discrimination occurs in the other direction. Thus the BellSouth simulation can be regarded as simulating a situation where there are systematic differences among the w/cs, with strong violations of parity in some w/cs and strong reverse effects in others.

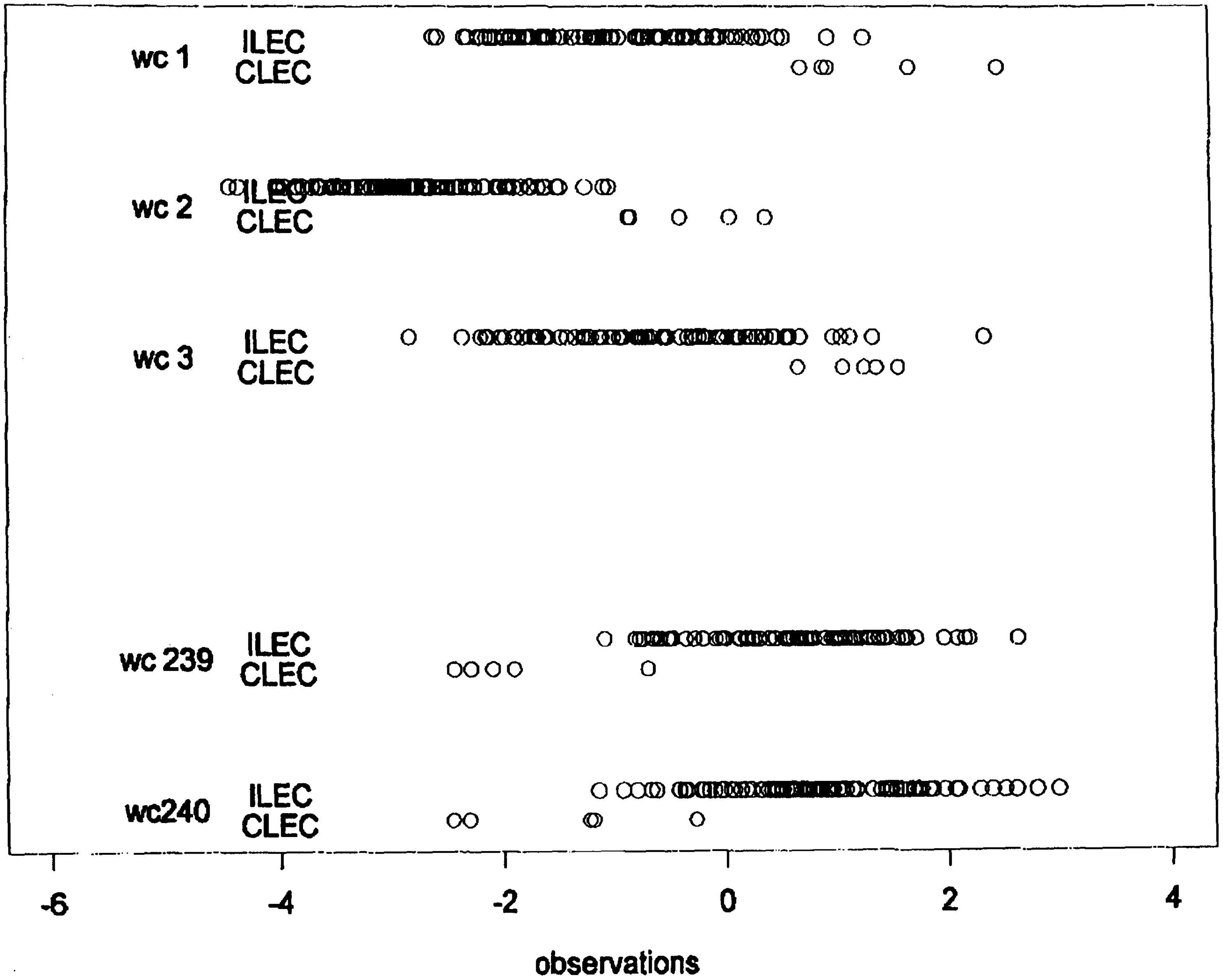
In the LCUG approach we compare the differences (ILEC sample mean – CLEC sample mean) (for each w/c) to measures of scale that are estimates of the within-w/c ILEC variance, which in these simulations is 0.65. Thus in this modified version of the simulation, in w/c #1 a highly significant Z-score (Z will be about –3.95) in each simulation run will be generated. In the BellSouth method, all of these within–w/c differences are regarded as being part of the overall random variation.

The simulation shows that the BellSouth method is completely insensitive to these effects. The BellSouth version of the LCUG method, since it uses within-w/c estimates of variance instead of the overall variance, gives an overall statistic that has the variance 1/0.65 = 1.539 instead of unity, so that the probability that it exceeds the 5% critical value is greater than 0.05. Thus this method does have a chance of detecting such systematic violations of parity, even though the average violation is zero. It is a positive attribute of the LCUG method that when these "correlations" are present, the method signals violation with probability greater than the size 0.05 of the test. The BellSouth method is completely insensitive to such systematic violations.


wc 1
ILEC
CLEC
wc 2
ILEC
CLEC
wc 3
ILEC
CLEC
wc 239
ILEC
CLEC
wc240
ILEC
CLEC
-6
-4
-2
0
2
4
observations


wc 1
ILEC
CLEC
wc 2
ILEC
CLEC
wc 3
ILEC
CLEC
wc 239
ILEC
CLEC
wc240
ILEC
CLEC
-6
-4
-2
0
2
4
observations


wc 1
ILEC
CLEC
wc 2
ILEC
CLEC
wc 3
ILEC
CLEC
wc 239
ILEC
CLEC
wc240
ILEC
CLEC
-6
-4
-2
0
2
4
observations



wc 1
ILEC
CLEC
wc 2
ILEC
CLEC
wc 3
ILEC
CLEC
wc 239
ILEC
CLEC
wc240
ILEC
CLEC
-6
-4
-2
0
2
4
observations

Here follows the S-plus function that generated the plots.

```
FCCplot_function(rho=.35,nILEC=100,nCLEC=5,.RS=.RS1){
.Random.seed <<- .RS
sqrtrho_sqrt(rho)
sqrt1mrho_sqrt(1-rho)
a_rnorm(240)*sqrtrho
b_rnorm(240)*sqrtrho
z1_matrix(rnorm(240*nILEC)*sqrt1mrho,nrow=240)
z2_matrix(rnorm(240*nCLEC)*sqrt1mrho,nrow=240)
x_z1+a
y_z2+b
# oamb_order(a-b)      # these three lines are made active to implement the `
# x_x[oamb,]           # second version of the program.
# y_y[oamb,]           #
plot(1:2,1:2,xlim=c(-6,4),ylim=c(1,27),xlab="observations",ylab="",yaxt="n",type="n")
points(x[1,],rep(27,nILEC))
points(y[1,],rep(26,nCLEC))
points(x[2,],rep(22,nILEC))
points(y[2,],rep(21,nCLEC))
points(x[3,],rep(17,nILEC))
points(y[3,],rep(16,nCLEC))
points(x[239,],rep(7,nILEC))
points(y[239,],rep(6,nCLEC))
points(x[240,],rep(2,nILEC))
points(y[240,],rep(1,nCLEC))
text(rep(-5,5),c(26.5,21.5,16.5,6.5,1.5),c("wc 1","wc 2","wc 3","wc 239","wc240"))
text(rep(-4,10),c(27,26,22,21,17,16,7,6,2,1),rep(c("ILEC","CLEC"),5))
}
> .RS1
 [1] 21 14 49 32 43  1 32 22 36 23 28  3
> FCCplot()                                       # first sheet
> .RS2_.Random.seed
> .RS2
 [1] 53 26 26 59 30  1 35 39 49 58 26  3
> FCCplot(.RS=.RS2)                               # second sheet
  .............   now edit the program to activate the sort ............
> FCCplot()                                       # third sheet
> FCCplot(.RS=.RS2)                               # fourth sheet
```

Exhibit 3

Appropriate Levels of Disaggregation For BellSouth Performance Measurements

Product Disaggregation

Resale		Unbundled Network Elements					Interconnection	
POTS	Specials	Loops	Ports		Transport	Combinations		
			Line	Trunk			Collocation	Trunks
Residence	Voice Grade Private Line	8db Analog	Analog	PRI	Dedicated DSO	Loop+Port+ Transport	Physical Caged Shared Caged Cageless	Common
Business	Digital DSO Private Line	2-wire digital	BRI	DID	Dedicated DS1	DS1 loop + office multiplexing	Virtual	Dedicated
Centrex/ Centrex-like	DS1	4-wire digital	DS1	Message	Dedicated DS3			
Analog PBX Trunks	DS3	ASDL			Dedicated >DS3			
DID trunks	Above DS3	HDSL						
ISDN BRI[1]		xDSL						
ISDN PRI[1]		All other unbundled loops						
All Other POTS-types								
All Other Specials								

Note 1. If treated as a designed service, the product detail may be more appropriately reflected within the "Resale Specials" category.

Appropriate Levels of Disaggregation
For BellSouth Performance Measurements

Activity Disaggregation

Pre-ordering (Data Exchange)	Ordering	Provisioning	Maintenance & Repair (Data Exchange)	Maintenance & Repair Tasks	Billing Usage	Billing Invoices
Address Verification	New Installation	Outside Dispatch	Trouble Entry	Outside Dispatch -No Service	End User	Resale
Telephone Number Requests	Change of Service Features	Central office Work (Frame or Equipment)	Trouble Status	Outside Dispatch -Degraded Service	Access	Unbundled Network Elements
Customer Account Information Requests	Disconnection	Software Only Work	Test Results	Central Office Work – No Service	Alternately Billed	Interconnection
Service/Product/ Facility Availability[1]	Inside Move	Disconnect	Trouble Cancellation	Central Office Work – Degraded Service		
Appointment/Due Date Scheduling[1]	Outside Move	Administrative	Rejection/Error	No Access to Premises or No Trouble Found		
Rejections/Errors	LSP-LSP Conversion without changes	No Access	Closure Notification	Administrative		
Loop Qualification Information	LSP-LSP Conversion with changes					
	Record Change Only					
	Standalone Directory Listing (DL)					
	Local Number Porting					
	Translation Disconnects					

Note 1. If transactions are differentiated, then performance should be separately tracked.

Appropriate Levels of Disaggregation
For BellSouth Performance Measurements

Activity Disaggregation (Continued)						
Pre-ordering	Ordering	Provisioning	Maintenance & Repair		Billing	
			(Data Exchange)	Tasks	Usage	Invoices
	Standalone Directory Assistance Listing (DA)					
	Standalone DL + DA					
	Other Orders					